Lecture # 11

Estimation

Estimation

Decision Theory

two major aspects

Estimation

Test of hypothesis

decision making under uncertain situation.

Introduction

The theory of statistical inference which has more recently become known as Decision Theory may be defined to be those methods by which one makes inferences or generalizations about a population based on information obtained from samples selected from the population. Therefore, decision theory is an important branch of statistics. In this branch we discuss the two major areas of decision theory, estimation and testing of hypothesis. This chapter deals with the estimation while the testing of hypothesis will be dealt with in the next book.

Estimation

Estimation means to estimate or predict the unknown value of a population parameter by the help of sample data. In other words, estimation refers to any procedure where sample information is used to estimate or predict the numerical value of some unknown population parameter. For example, if a candidate for public office may wish to estimate the true proportion of voters favouring him by obtaining the opinions from a random sample of 200 eligible voters. The fraction of

population prop...

Methods of obtaining point Estimators

A point estimator of a parameter can be obtained by several methods but we shall consider the following three methods

(i) Method of Maximum Likelihood. MLE

(ii) Method of Moments MME

(iii) Method of Least Squares. ML

Method of Maximum Likelihood

MML

Start To illustrate the method we assume that the population has a density function which contains a population parameter say θ, which is to estimated by a certain statistic. Thus the density function can be denoted by $f(x;\theta)$. PDF

In order to define maximum likelihood estimator, we shall define first the likelihood function:

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a probability density function $f(x;\theta)$ where θ is an unknown parameter. Therefore the joint probability density function for a sample of n independent observations is

$$f(x_1, x_2, \ldots, x_n;\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \ldots f(x_n;\theta)$$

$$= \prod_{i=1}^{n} f(x_i;\theta)$$

This joint p.d.f. regarded as a function of θ and is called the likelihood function of the sample and is denoted by $L(\theta)$ or sometimes by L.

i.e. $\quad L.F. = L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

The method of maximum likelihood states that the statistic $\hat{\theta}$ is called the estimator of θ if $\hat{\theta}$ maximizes likelihood function. Thus $\hat{\theta}$ is a solution, if any, of Log apply | differentiate

$$\frac{dL(\hat{\theta})}{d\theta} = 0 \text{ or sometime } \frac{d \log L(\theta)}{d\theta} = 0$$

$$\text{with } \frac{d^2 L(\theta)}{d\theta^2} < 0 \text{ or sometimes } \frac{d^2 \log L(\theta)}{d\theta^2} < 0$$

Therefore the value of θ, for which the likelihood function is maximum, is the maximum likelihood estimator of θ.

Find the maximum likelihood estimator of the parameter λ in the poisson distribution. $f(x;\lambda)$

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x = 0, 1, 2, 3, \ldots, \infty$$

Solution :

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample drawn from a poisson distribution.

therefore

$$P(X_1) = \frac{e^{-\lambda} \lambda^{X_1}}{X_1!}$$

$$P(X_2) = \frac{e^{-\lambda} \lambda^{X_2}}{X_2!}$$

$$P(X_n) = \frac{e^{-\lambda} \lambda^{X_n}}{X_n!}$$

Log Rule

$\log(a/b) = \log a - \log b$

$\log e = 1$

$\log(a \times b) = \log a + \log b$

$\log a^n = n \log a$

Then the likelihood function defined as

$$L = P(X_1) \cdot P(X_2) \ldots P(X_n)$$

$$= \frac{e^{-\lambda} \lambda^{X_1}}{X_1!} \cdot \frac{e^{-\lambda} \lambda^{X_2}}{X_2!} \cdots \frac{e^{-\lambda} \lambda^{X_n}}{X_n!}$$

$$= \frac{e^{-n\lambda} \cdot \lambda^{X_1 + X_2 + \ldots + X_n}}{X_1! \cdot X_2! \ldots X_n!} = \frac{e^{-n\lambda} \cdot \lambda^{n\bar{X}}}{\prod_{i=1}^{n} X_i!}$$

$\bar{x} = \sum x_i / n$, $n\bar{x} = \sum x_i$

Log Apply :-

and $\log_e L = -n\lambda + n\bar{X} \log_e \lambda - \log_e \prod_{i=1}^{n} X_i!$

Apply derivative :-

$$\text{Now} \quad \frac{d \log L}{d\lambda} = -n + \frac{n\bar{X}}{\lambda} - 0 = 0$$

$$-n + \frac{n\bar{X}}{\lambda} = 0$$

$$-1 + \frac{\bar{X}}{\lambda} = 0$$

$$\boxed{\hat{\lambda} = \bar{X}}$$

$\frac{d}{d\lambda} = -n \frac{d\lambda}{d\lambda}$

$\frac{d}{d\lambda} = n\bar{x} \frac{d}{d\lambda} \log \lambda$

$= n\bar{x} \frac{1}{\lambda}$

$\frac{d}{d\lambda} = \log \pi x_i$

binomial

Point Estimators

- Find ,MLE of the parameter p in the binomial distributions

$$P(x;n,p) = {}^nC_x p^x q^{n-x}, \quad x = 0,1,2,\dots,n$$

Solution :

Let $X_1, X_2, \dots X_m$ be a random sample drawn from binomial distribution.

then the likelihood function is ;

$$L = P(X_1;n,p).P(X_2;n,p)\dots P(X_m;n,p)$$

$$= {}^nC_{X_1}.p^{X_1}.q^{n-X_1}.{}^nC_{X_2}p^{X_2}.q^{n-X_2}\dots{}^nC_{X_m}p^{X_m}.q^{n-X_m}$$

$$= \prod_{i=1}^{m}{}^nC_{X_i}.p^{\sum X_i}.(1-p)^{\sum(n-X_i)} \quad \text{where } q = 1-p$$

$$= \prod_{i=1}^{m}{}^nC_{X_i}\,p^{m\overline{X}}.(1-p)^{nm-m\overline{X}}$$

$$\log_e L = \log_e \prod_{i=1}^{m}{}^nC_{X_i}\,p^{m\overline{X}}.(1-p)^{nm-m\overline{X}}$$

Applying log $\log_e L = \log_e \prod_{i=1}^{m}{}^nC_{X_i} + m\overline{X}\log_e p + (nm - m\overline{x})\log_e(1-p)$

App diff : $\dfrac{d\log_e L}{dp} = 0 + \dfrac{m\overline{X}}{p} + \dfrac{nm-m\overline{X}}{1-p}(-1) = 0$

$$\frac{m\overline{X}}{p} = \frac{nm - m\overline{X}}{1-p}$$

$\dfrac{d}{dp}\log_e p = \dfrac{1}{p}$

$$\frac{\overline{X}}{p} = \frac{n-\overline{X}}{1-p}$$

$$\overline{X}(1-p) = p(n-\overline{X})$$

$$\overline{X} - p\overline{X} = np - p\overline{X}$$

$$\boxed{\overline{X} = np}$$

Then $\hat{p} = \dfrac{\overline{X}}{n}$ where $\overline{X} = \dfrac{\sum_{i=1}^{m} x_i}{m}$

Hence $\hat{p} = \dfrac{\overline{X}}{n}$ is the MLE of parameter p

... parameter p.

Find M.L.E of β in the exponential distribution.

$$f(x;\beta) = \frac{1}{\beta}e^{-x/\beta}, \quad 0 \le x < \infty$$

Solution:

Take a random sample of n independent values,

$$X_1, X_2, \dots X_n$$

then the likelihood function is

$$L = f(X_1;\beta).f(X_2;\beta)\dots f(X_n;\beta)$$

$$= \frac{1}{\beta}e^{-X_1/\beta}.\frac{1}{\beta}e^{-X_2/\beta}\dots\frac{1}{\beta}e^{-X_n/\beta}$$

$\dfrac{e^{-n\overline{X}/\beta}}{\beta^n} = \dfrac{1}{\beta^n}.e^{-n\overline{X}/\beta}$ $\sum x_i = n\overline{X}$

$$\log_e L = \frac{-n\overline{X}}{\beta} - n\log_e\beta$$

$$\frac{d\log_e L}{d\beta} = \frac{n\overline{X}}{\beta^2} - \frac{n}{\beta} = 0$$

$$\frac{n\overline{X}}{\beta^2} = \frac{n}{\beta}$$

$$\frac{\overline{X}}{\beta} = 1$$

$$\boxed{\hat{\beta} = \overline{X}}$$

Method of Moments MME ·

The method of moments consists of equating the first k-moments of a population to the corresponding moments of a sample. Then solving these k-equations for the k-unknown parameters. Sample moment = popN moment

Suppose μ_r' and m_r' are the rth moments about origin of the population and sample respectively then the solution of the equations $m_r' = \mu_r'$ ($r = 1, 2, \ldots k$) yields the results of the unknown parameters.

class ·

Use the method of moments to estimate the parameter θ in the following uniform distribution.

$$f(x;\theta) = \frac{1}{\theta}, \qquad 0 < x < \theta \qquad f(x) = \frac{1}{b-a}$$

$\theta - 0 = \theta$

Solution :

Since there is only one parameter to be estimated, therefore we need one equation

For popN

then $\mu_1' = E(X) = \int_0^\theta x \cdot \frac{1}{\theta} \cdot dx = \boxed{\frac{\theta}{2}}$

$E(X) = \int x f(x)\,dx = \int_0^\theta x \frac{1}{\theta}\,dx = \frac{1}{\theta}\left|\frac{x^2}{2}\right|_0^\theta = \frac{1}{2\theta}\left|\theta^2 - 0^2\right| = \frac{\theta^2}{2\theta} = \boxed{\frac{\theta}{2}}$

Take a random sample of n independent values then the corresponding first sample moment is

For sample — moments about origin

$$m_1' = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

$\mu_1' = \frac{\sum x}{n} = \bar{x}$

Now equating m_1' and μ_1' for pop def.

$$\mu_1' = m_1'$$

$$\boxed{\frac{\theta}{2} = \overline{X}}$$

$$\hat{\theta} = 2\overline{X}$$

The density function of a normal distribution is given as follows.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Find the estimators for μ and σ^2 by the method of moments

Solution :

Since two parameters μ and σ^2 to be estimated therefore we need two equations for solutions, these equations are

$$m_1' = \mu_1' \quad \text{and} \quad m_2' = \mu_2'$$

since $\boxed{\mu_1' = \mu \quad \text{and} \quad \mu_2' = \mu^2 + \sigma^2}$

normal distribution

$\mu_1' = \mu$

Now the sample moments for a sample of size n are ; $\mu_2' = \mu^2 + \sigma^2$

$\mu_3' = 3\sigma^2\mu + \mu^3$

moments about origin

$$m_1' = \frac{\Sigma X_i}{n} \quad \text{and} \quad m_2' = \frac{\Sigma X_i^2}{n}$$

since $\quad \mu_1' = m_1'$

then $\quad \hat{\mu} = \overline{X}$

and $\quad \mu_2' = m_2'$

$$\mu^2 + \sigma^2 = \frac{\Sigma X_i^2}{n}$$

$$\sigma^2 = \frac{\Sigma x_i^2}{n} - \mu^2$$

$$\hat{\sigma}^2 = \frac{\Sigma X_i^2}{n} - \overline{X}^2 \quad \text{since } \hat{\mu} = \overline{X}$$

Hence the estimator of μ is $\overline{X}$ and the estimator of σ^2 is

$$\frac{\Sigma X_i^2}{n} - \overline{X}^2 \quad \text{or} \quad \frac{\Sigma (X_i - \overline{X})^2}{n}$$

Interval Estimators

Method of Least Squares: MLE

According to this method, an estimator found by minimizing the sum of squares of deviations of the sample values from some function that has been hypothesized as a fit for the data, is called the least squares estimator.

Interval Estimation:

Any point estimate has the limitation that it does not provide information about the precision of the estimate i.e. about the magnitude of the error due to sampling. In other words, we can say point estimates are not good estimates of population parameters because these estimates fail to throw light on how close we can expect such as estimate to be to the population parameter we wish to estimate. Thus, we cannot associate a probability statement with point estimates. We therefore try to determine two values, instead of one point estimate within which the true value of the parameter is expected to lie with a certain degree of confidence i.e. $(1 - \alpha)$ 100%. The limits which contain a population parameter with a given degree of confidence are called the confidence limits. (lower and upper confidence limits). The interval between these limits is called the confidence interval or interval estimate and the whole process is known as interval estimation or estimation by confidence interval. It is important to note, if we determine a 95% confidence interval, we understand that the probability that the interval contains the true parameter is 0.95 or in other words out of 100 possible intervals 95 of the intervals are certain to contain the true parameter.

• Findout the popn mean ; when given popn Variance :-

1 point estimation :

• 1000 Students (popn) university

• 80 Students (sample)

want to find age of 80 student ;

1 20
2 21
3 22
4 23
5 24
6 27
7 20
⋮
80 20

$\bar{x}$ 20.5 age

$\boxed{E(\bar{x}) = \mu}$ from sampling dist .

• but; $E(\bar{x}) \notin$ ~~not~~ all sample Value .

my sample 80 ≠ 1000 popn sample .

20.5 ≠ 20.5 — Point estimation

Not true mean value it just approximate (Estimation)

2 Interval Estimation

range define :-

$19 < \mu < 23$

may be the mean and

Confidence Interval

↳ also called
↳ confidence limit
↳ interval estimation
↳ whole process is called interval estimation

• 90% Sure that mean lie this range .

$19 < \mu < 25$

Interval estimation

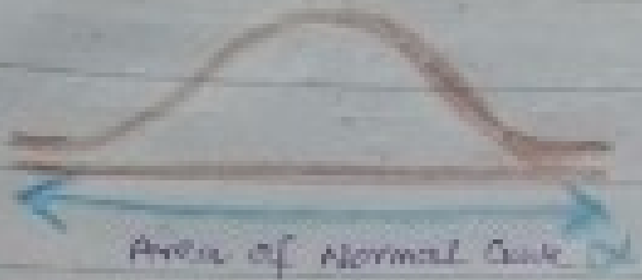
Formula FOR μ :-

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{n}$$

n-2

Confidence level:-

90%, 95%, 99% → normal use 96, 98 also use

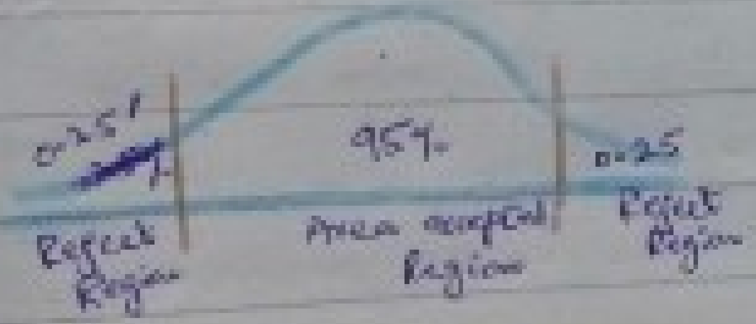


if:

$\alpha = 5\% \Rightarrow 0.05$ coefficient

$100(1-\alpha)$ = confidence ~~level~~

$1 - 0.05 = 0.95 \times 100 = 95\%$



95% = accepted

0.5 = Rejected (may be error).

↓

$\alpha/2 = 0.25$ Left/Right

$$P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

therefore a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

or it may be written as

$$\boxed{\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

where $1-\alpha$ is called the confidence coefficient.

Confidence Interval for Population mean μ

Case (i) • When population is normal

• When sample size is large i.e. $n \geq 30$

• When standard deviation of population σ is known

Then the procedure to determine a $100(1-\alpha)\%$ confidence interval for population mean μ is as under.

Example 9

An electrical firm manufactures T.V. picture tubes that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a random sample of 30 tubes has an average lie of 780 hours, find a 96% confidence interval for the population mean of all tubes produced by this firm.

Solution :

Since $n = 30$, $\overline{X} = 780$ and $\sigma = 40$

Also confidence coefficient $1-\alpha = 0.96$. Therefore

$\boxed{\alpha = 0.04}$ and $\boxed{\alpha/2 = 0.02}$.

Then $100(1-\alpha)\%$ C.I. for μ is given by

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Thus 95% C.I. for μ is;

$$780 - Z_{0.02}\frac{(40)}{\sqrt{30}} < \mu < 780 + Z_{0.02}\frac{(40)}{\sqrt{30}}$$

$$780 - \frac{(2.06)(40)}{\sqrt{30}} < \mu < 780 + \frac{(2.06)(40)}{\sqrt{30}}$$

$$780 - 15.044 < \mu < 780 + 15.044$$

$$764.95 < \mu < 795.044$$

or $\boxed{765 < \mu < 795}$

Case (ii) • when population is normal

• when $n < 30$

• when σ is known

Case (ii)

- when population is normal
- when $n < 30$
- when σ is known

Then a 100 $(1-\alpha)$ % confidence interval for population mean μ is given by ;

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

A random sample of size 6 is taken from a normal population with a known standard deviation $\sigma = 2.50$. If the mean of the sample is 8.00, find 95% confidence interval for the population mean μ.

Solution :

Since standard deviation of population is known and also the population is normal, therefore a $100(1-\alpha)\%$ C.I. for μ is ;

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

and since $\bar{X} = 8.00$, $n = 6$, $\sigma = 2.50$, $1-\alpha = 0.95$ and $\alpha = 0.05$,

Then 95% C.I. for μ is ;

$$8.00 - Z_{0.025}\frac{(2.50)}{\sqrt{6}} < \mu < 8.00 + Z_{0.025}\frac{(2.50)}{\sqrt{6}}$$

$$8.00 - \frac{(1.96)(2.50)}{\sqrt{6}} < \mu < 8.00 + \frac{(1.96)(2.50)}{\sqrt{6}}$$

$$8.00 - 2.00 < \mu < 8.00 + 2.00$$

or $\quad 6 < \mu < 10$

Case (iii)

- When population is normal
- when population standard deviation σ is not known but sample standard deviation s is known
- and when $n \geq 30$.

Then a $100(1-\alpha)\%$ C.I. for population mean μ is given by

$$\overline{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Example 11

A random sample of size $n = 64$ is taken from a normal population with unknown mean and variance. The sample mean is 120 and standard deviation 5. Set up a 90% confidence interval for the population mean μ.

Solution :

Since σ is unknown but $n \geq 30$, therefore we replace σ by the sample standard deviation s.

Hence 90% confidence interval for μ is computed as :

since $\overline{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$

then $120 - Z_{0.05}\frac{(5)}{\sqrt{64}} < \mu < 120 + Z_{0.05}\frac{(5)}{\sqrt{64}}$

and $120 - \frac{(1.65)(5)}{\sqrt{64}} < \mu < 120 + (1.65)\frac{(5)}{\sqrt{64}}$

$120 - 1.03 < \mu < 120 + 1.03$

$118.97 < \mu < 121.03$

Case (iv)
* when population is non-normal
* when σ is known and $n \geq 30$

OR

* when σ is unknown and $n \geq 50$

Then a $100(1-\alpha)\%$ C.I. for population mean μ for known σ is given by;

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

and also a $100(1-\alpha)\%$ C.I. for population mean μ for unknown σ is given by

$$\overline{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Where $s \simeq \sigma$ for large n.

EXAMPLE # 02

A random sample of 100 observations from a population known to be non-normal yielded the sample result; $\overline{X} = 14500$ and $s = 2400$. Find an approximate 99% confidence interval for μ.

Solution :

Since size of sample is large enough i.e. $n = 100 > 30$ therefore the sampling distribution of $\overline{X}$ is approximately normal with mean μ and standard deviation $\frac{s}{\sqrt{n}}$

Hence a $100(1-\alpha)\%$ C.I. for μ is

$$\overline{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $\overline{X} = 14500$, $n = 100$, $s = 2400$

therefore 99% C.I. for μ is;

$$14500 - Z_{0.005}\cdot\frac{2400}{\sqrt{100}} < \mu < 14500 + Z_{0.005}\cdot\frac{2400}{\sqrt{100}}$$

$$14500 - \frac{(2.58)(2400)}{\sqrt{100}} < \mu < 14500 + \frac{(2.58)(2400)}{\sqrt{100}}$$

$$13880.8 < \mu < 15119.2$$

EXAMPLE # 01

A sample of 35 observations is taken from a non-normal population with unknown mean μ and known standard deviation $\sigma = 8$. If the mean of the sample is 17.2 Find 90% confidence interval for the population mean μ.

Solution :

Since the sample size is fairly large ($n = 35 > 30$) and since the population standard deviation is known. Therefore a $100(1-\alpha)\%$ confidence interval for population mean μ is given by.

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $\overline{X} = 17.2$, $\sigma = 8$, $n = 35$ and $1 - \alpha = 0.90$

then $\alpha = 0.10$.

Therefore 90% C.I. for μ is;

$$17.2 - Z_{0.05}\cdot\frac{8}{\sqrt{35}} < \mu < 17.2 + Z_{0.05}\cdot\frac{8}{\sqrt{35}}$$

or $$17.2 - \frac{(1.645)(8)}{\sqrt{35}} < \mu < 17.2 + \frac{(1.645)(8)}{\sqrt{35}}$$

or $$15.0 < \mu < 19.4$$

Case V (case of t-distribution)

- when population is normal
- when sample size is small i.e. $n < 30$ and
- when standard deviation of population σ is unknown (but sample standard deviation s is known)

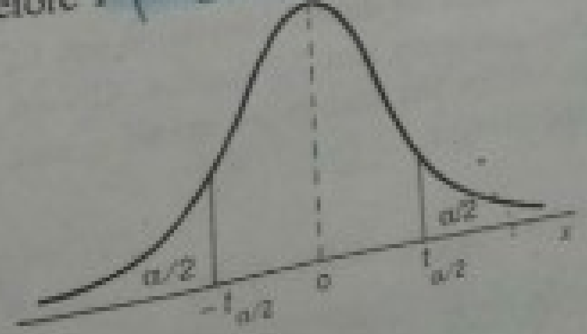
Then the confidence interval for population mean μ is based on t–distribution. The t–distribution is defined as ;

Let $X_1, X_2, X_3 \ldots\ldots X_n$ be a random sample drawn from a normal population with mean μ and unknown variance σ^2, then the sampling distribution of the statistic $t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$ is called t–distribution with (n–1) degrees of freedom.

where $s^2 = \frac{\Sigma (X - \overline{X})^2}{n-1}$ is the unbiased estimate of σ^2,

$\overline{X}$ is the sample mean or point estimate of μ and n is the sample size provided n < 30.

Therefore $P\left(-t_{\alpha/2} < t < t_{\alpha/2}\right) = 1 - \alpha$



and since $t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$

therefore $P\left(-t_{\alpha/2} < \frac{\overline{X} - \mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$

and $P\left(\overline{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$

Then for a particular random sample of size n < 30, the 100 (1 – α)% C.I. for μ is given by

$$\overline{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

or $\boxed{\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}}$

Where $\overline{X}$ and s are the sample mean and standard deviation of a random sample of size n < 30 from a normal population and $t_{\alpha/2}$ is the value of t–distribution with (n–1) degrees of freedom at the level of significance $\alpha/2$ on each side of the tail of t–distribution.

Degrees of Freedom

The statistical concept of degrees of freedom is one of the most difficult for begining students because of its many possible interpretation. The general expression for degrees of freedom is (n–k), where n is the number of observations and k is the number of constants that must be calculated from the sample data to estimate the variance of the sampling distribution.

If a random sample of 15 measurements of the breaking strength of cotton threads, the mean breaking strength was found to be 7 ounces and the standard deviation was 1.5 ~~ounces and the standard deviation was 1.5 ounces.~~ Obtain a 90% confidence interval for the true mean breaking strength of cotton threads of this type.

Solution :

A $100(1-\alpha)\%$ C.I. for μ is

$$\overline{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{with} \quad \text{d.f.} = \nu = n-1$$

where $\overline{X} = 7$, $s = 1.5$, $n = 15$ and $1-\alpha = 0.90$

then 90% C.I. for μ is

$$7 \pm t_{0.05} \frac{(1.5)}{\sqrt{15}}$$

$$7 \pm \frac{(1.761)(1.5)}{\sqrt{15}} \quad \text{since} \quad t_{0.05,\nu=14} = 1.761$$

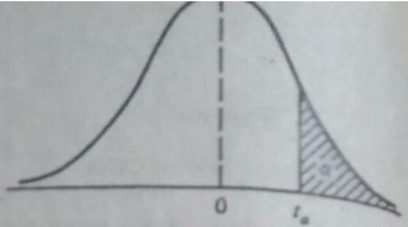
$$7 \pm 0.68$$

$$6.32,\ 7.68$$

or $\quad 6.32 < \mu < 7.68$

Table–2

Critical Values of the t Distribution



ν	α = 0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576

① Confidence Interval for popn mean 'μ'

① popn Normal
sample size $n \geq 30$
σ is known

② popn Normal
$n < 30$
σ is know

③ popn Normal
σ is unknow but sample SD
s is know
$n \geq 30$

formula: $\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\, \sigma/\sqrt{n}$ (1-4 cases)

④ popn non-normal
σ known
$n \geq 30$
OR
σ unknown $n \geq 50$

⑤ popn is Normal
sample size $n < 30$
σ unknow
s known.

$$\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

② Confidence Interval for the difference b/w two popn mean ($\mu_1 - \mu_2$)

① two popn Normal
- $\sigma_1\ \sigma_2$ known
- $n_1, n_2 \geq 30$ OR
$n_1, n_2 < 30$

② two popn Normal
$\sigma_1\ \sigma_2$ unknown
sample size $n_1 n_2 \geq 30$

③ popn non-Normal
$\sigma_1\ \sigma_2$ know or unknown
$n_1, n_2 \geq 30$

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ (1-3 cases)

④ popn Normal
- $n_1, n_2 < 30$
- $\sigma_1 = \sigma_2$ unknown

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

⑤ popn normal.
$n_1 < 30$ $n_2 < 30$
$\sigma_1 \neq \sigma_2$ unknown.

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

③ Confidence interval for the diff b/w two popn mean (for paired observation)

$\mu_D = \mu_1 - \mu_2$

$$\bar{d} - t_{\alpha/2} \frac{S_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2} \frac{S_d}{\sqrt{n}}$$

∴ $\bar{d} = \frac{\sum d_i}{n}$ $S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$

④ Confidence Interval for population proportion 'p'

$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

- $\hat{p} = x/n$ (Proportion of success)
- $n \geq 30$
- $\hat{q} = 1 - \hat{p}$

⑤ Confidence Interval for difference of two popn proportion ($P_1 - P_2$)

- $n_1, n_2 \geq 30$

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

- $\hat{p}_1 = \frac{x_1}{n_1}$
$\hat{p}_2 = \frac{x_2}{n_2}$
$\hat{q}_1 = 1 - \hat{p}_1$; $\hat{q}_2 = 1 - \hat{p}_2$

⑥ Confidence Interval for popn Variance σ^2

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \qquad s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$ σ = popn var

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

⑦ Confidence Interval for population Correlation coefficient 'ρ'

$$\frac{e^{2w_1} - 1}{e^{2w_1} + 1} < \rho < \frac{e^{2w_2} - 1}{e^{2w_2} + 1}$$

∴ $w_1 = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) - \frac{Z_{\alpha/2}}{\sqrt{n-3}}$

$w_2 = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) + \frac{Z_{\alpha/2}}{\sqrt{n-3}}$